Linear integrated location-inventory models for service parts logistics network design

Fatma Gzara

Department of Management Sciences, University of Waterloo

Eissa Nematollahi

Schulich School of Engineering, University of Calgary

Abdullah Dasci,

School of Management, Sabanci University

January, 2013

Abstract

We present two integrated network design and inventory control problems in service-parts logistics systems. Such models are complicated due to demand uncertainty and highly nonlinear time-based service level constraints. Exploiting unique properties of the nonlinear constraints, we provide an equivalent linear formulation under part-warehouse service requirements, and an approximate linear formulation under part service requirements. Computational results indicate the superiority of our approach over existing approaches in the literature.

Keywords: service-parts logistics, logistics network design, facility location, inventory management, time-based service requirements.

Linear location-inventory models for service parts logistics network design

Abstract

We present two integrated network design and inventory control problems in service-parts logistics systems. Such models are complicated due to demand uncertainty and highly nonlinear time-based service level constraints. Exploiting unique properties of the nonlinear constraints, we provide an equivalent linear formulation under part-warehouse service requirements, and an approximate linear formulation under part service requirements. Computational results indicate the superiority of our approach over existing approaches in the literature.

Keywords: service-parts logistics, logistics network design, facility location, inventory management, time-based service requirements.

1 Introduction

After-sales services include inspection, repair, and provisioning of service parts, and represent a lucrative part of the overall operations in such industries as aerospace, automotive, and high-technology. For example, service parts management is an integral part of original equipment manufacturers' after-sales operations that may account for 40-50% of total profits [1, 2]. After-sales service is particularly critical for the success of high-technology equipment manufacturers whose clients rely on the availability of equipment for their daily operations and therefore, expect high quality and fast service. On the one hand, advances in information technology led to increased practice of after-sales contracts by allowing clients to monitor the terms of the contracts and design performance-based incentive schemes [3]. On the other hand, global competition reduced profit margins and forced firms to differentiate themselves by the after-sales services they offer.

To provide high quality service, a responsive *service-parts logistics* (SPL) system is needed. In such a system each customer has a service center (or, warehouse hereafter) within its close proximity to provide the necessary parts when needed. Responsiveness, however, comes at a significant cost as a typical SPL system may consist of tens or hundreds of warehouses and hundreds of stock-keeping units. Therefore, it is important that firms design their SPL systems rationally to control costs while providing the desired level of responsiveness.

There are two important sets of decisions that determine the responsiveness of a SPL system. The first set of decisions concerns location and allocation and determine the proximity of warehouses to customers. These decisions are strategic in nature and are part of *logistics network design* (LND). The second set of decisions concerns inventory and are operational/tactical in nature. They determine the responsiveness of the system in terms of stock availability. Traditionally, the two sets of decisions are made at different levels of management and are generally analyzed separately. However, integrating LND decisions with inventory decisions leads to better network design and potentially better inventory policies, enabling firms to offer similar or improved responsiveness at lower cost.

In this paper, we study a series of integrated models for designing SPL systems that model

LND decisions and inventory control policies at the warehouses to minimize the total cost of warehouse opening, transportation and inventory holding. The models belong to the general class of integrated location-inventory models. Both LND and inventory decisions, in isolation, are studied extensively in the Operations Management literature. There is also growing literature on integrated location-inventory models. However, our models have a number of differentiating features. Primarily, we consider only service parts that are expensive, reliable, and experience low and random demand, for which base-stock $(s-1, s)$ policies are known to be optimal. Most of the past location-inventory literature consider generic items and usually use the economic-order-quantity reorder-point (EOQ-ROP) approach as an approximation to (s, S) policy. Most importantly, we consider time-based service level requirements whereby the manufacturer guarantees delivery of service parts within a specified time window at a specified percent of the time. To the best of our knowledge, the first attempt to model and solve a problem of this class is given by [4]. Their model is equivalent to one of the models we study here and will be elaborated on later.

We consider a two-echelon SPL system with one distribution center, a number of potential warehouse locations, and a set of customers dispersed across a geographical market. The LND part is modeled as a multi-product fixed-charge facility location problem, and involves warehouse opening and customer allocation decisions. Figure 1 shows an SPL system with three open warehouses and ten customers. The inventory decisions depend on the underlying SPL system, and concern setting inventory levels to satisfy customer demand as prescribed in the service agreements.

Stocking decisions depend on how the service level measures are defined and incorporated into inventory management. In practice, after-sales service contracts may be quite complex and may involve basket products, composite or indirect measures, incentive or penalty schemes, etc. Here we consider two time-based service level measures. The *part-specific service level* measure is an aggregate measure where the manufacturer makes a system-wide target service level for each part, which has to be achieved by the warehouses collectively. A typical requirement may read: "*A part's overall demand is satisfied* 90% *of the time within a 4 hour time window*

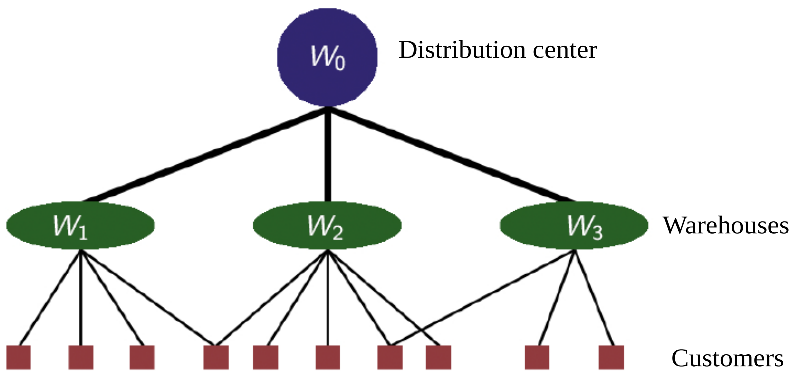

Figure 1: A two-echelon service-parts logistics system.

across all customers and warehouses." In the *part-warehouse service level*, the manufacturer exogenously sets target service levels for each warehouse and part, then determines the inventory stocking decisions at the warehouses. A typical service level requirement in this case may read: "*At least 90% of a part's demand assigned to a warehouse must be satisfied within 4 hours.*" The part service level is a more aggregate measure, whereas the part-warehouse service level gives an upper bound when the same target level is set across all warehouses.

Location and allocation decisions that take inventory decisions into account are known to be different from those that do not [5]. Nevertheless, inventory considerations even under deterministic assumptions present nonlinearities that may be quite challenging [6, 7]. A number of integrated models are studied by [8, 9, 10] and [11, 12] who introduce joint location-inventory models and exploit various heuristics or approximations to solve the resulting nonlinear optimization problems. Nonlinear location models with risk pooling are studied in [13, 14] and [15]. An extension that considers customer service levels is presented in [16]. All of these works are different from ours, as they consider generic products which have relatively high demand. We, on the other hand, deal with service parts that generally have very low demand. This difference naturally leads to different inventory control policies and substantially different models.

Integrated models that consider LND and inventory decisions for service parts have been the subject of research only recently. To the best of our knowledge, [4] was the first to propose a mixed-integer nonlinear formulation to model time-based service level requirements with multiple parts and backorders. The authors present a linear approximation based on the discretization of the fill rate functions at the warehouses. The approximation however does not guarantee feasibility with respect to service level requirements and may find solutions that violate target service level. In a post-processing step, they solve a pure inventory problem to boost stock levels. The work in [17] studies a single-part case with lost sales and proposes a mixed-integer nonlinear formulation. The authors develop an outer approximation to find a heuristic solution and a lower and an upper bound. The lower bound is found by ignoring the customers outside the time-window, whereas the upper bound is found by solving a series of approximate models with relaxed service requirements. The authors in [18] describe a network design and inventory control project undertaken at Applied Materials, the largest semiconductor equipment producer in the world. They consider both service parts and generic products and propose a piece-wise linear underestimation for the inventory costs that is updated progressively. This is then solved in a proprietary solution platform with two underestimation pieces to obtain an approximate solution. Since they do not provide any detailed results, we could not make a comparison with their model. The work in [19] studies the expected behavior of an SPL system and presents a mixed-integer programming model to minimize the expected cost of inventory and backorder subject to an upper bound limit on the expected wait time.

In this work, we propose two mixed-integer nonlinear formulations for integrated location and inventory decisions under part-warehouse and part service level requirements. The model under part service level is similar to that studied in [4], while both models are substantially different from that in [17]. There are three differentiating features. We consider multiple parts and backorders, and require that all customers are served from a warehouse within the time-window. The authors in [17] consider single part and lost sales, and allow customers to be served from a warehouse outside the time-window. Since the fill rate functions under backorders and lost sales are substantially different [20], the corresponding models and approximations are not directly comparable. Furthermore, it is straightforward to treat the single-part case using a multiple-part model but extending a single-part model to multiple parts presents serious challenge.

The integrated models are extremely difficult to solve given the large number of nonlinearities introduced by service level constraints. We develop a novel approach to approximate these nonlinearities to lead to linear reformulations that are effectively solved by a commercial optimizer in a modest computing environment. We replace the fill-rate function by its inverse in the part-warehouse specific case, and discretize a special function of the fill rate in the part specific case. A significant contribution is that in the case of part-warehouse service levels, the linear formulation is equivalent to the nonlinear model. Moreover, the part-warehouse specific case provides an upper bound on the more difficult part-specific case when the same target service level is used across all warehouses. The results suggest that our approach is more effective than those reported in the literature both in terms of solution quality and computational time. The proposed formulations are shown to be effective in solving larger instances.

The rest of the paper is organized as follows. In Section 2, we present models for part-warehouse, and part specific cases and develop their linear counterparts. In Section 3, we report on computational experiments. Finally, in Section 4 we provide concluding remarks and potential future research directions.

2 Linear location-inventory models

In this section, we present models and analysis for the part-warehouse and part service levels. We first specify the assumptions common to both models and to related literature.

- The distribution center (DC) has an unlimited capacity.
- Warehouses use a continuous review and $(s-1, s)$ replenishment policy with backordering.
- Customer demands arrive one at a time according to an independent Poisson distribution.
- Parts are shipped directly to corresponding customers without consolidation or bundling.

- Customers are served on a first-come first-serve basis regardless of their location.

- Lateral transshipments are not allowed among warehouses.

Let $\mathcal{I}_m = \{1, 2, \ldots, m\}$, $\mathcal{I}_n = \{1, 2, \ldots, n\}$, and $\mathcal{I}_p = \{1, 2, \ldots, p\}$ denote the index sets for potential warehouse locations, customers, and parts, respectively. Designing and operating an SPL system involves incurring an annual fixed cost f_i for each open warehouse i, a unit holding cost h_{ik} for part k at warehouse i, and a unit transportation cost c_{ijk} for part k from warehouse i to customer j. The service time, which includes the transportation time, from warehouse i to customer j for part k is denoted by τ_{ijk}. The lead time for part k from the distribution center to warehouse i is t_{ik}. The demand of customer j for part k is a random variable $\widetilde{d}_{jk}$ with Poisson distribution and mean annual demand d_{jk}.

The models use three main sets of decision variables. Variable $x_{ijk} \in [0, 1]$ represents the long-run fraction of customer j's demand for part k fulfilled from warehouse i. Binary variable y_i takes value 1 if warehouse i is open, and 0 otherwise. Integer variable $s_{ik} \in \mathcal{I}_{S_{\max}} = \{1, 2, \ldots, S_{\max}\}$ represents the base stock level of part k at warehouse i, where $S_{\max}$ is the stocking capacity set a priori on the stock level of a part at a warehouse. The notation is summarized in Table 1.

Table 1: Summary of notation.

f_i	Annual fixed location cost for warehouse i.
c_{ijk}	Unit transportation cost from warehouse i to customer j for part k.
h_{ik}	Unit holding cost for part k at warehouse i.
τ_{ijk}	Transportation time from warehouse i to customer j for part k.
t_{ik}	Lead time from DC to warehouse i for part k.
$\widetilde{d}_{jk}$	Random customer j demand for part k with mean annual demand d_{jk}.
$S_{\max}$	Upper bound on stock level.
y_i	Binary location variable.
x_{ijk}	Long-run fraction of customer j demand for part k served from warehouse i.
s_{ik}	Stock level of part k at warehouse i.
w_k	Time window for part k.

We require that there exists at least one warehouse within the time window w_k of each customer for each part, i.e.,

$$\tau_{ijk} \le w_k, \qquad \forall j \in \mathcal{I}_n \; \forall k \in \mathcal{I}_p \; \exists i \in \mathcal{I}_m. \tag{1}$$

Note that without condition (1), the problem may be infeasible since not every customer can be serviced by a warehouse within the time window. On the practical side, it is not possible to promise a customer delivery within a specified time window when there is no open warehouse within that window. To overcome this, we require that customers are allocated to warehouses within the time window. Therefore, in all cases we have

$$\sum_{i \in \mathcal{I}_m :\, \tau_{ijk} \le w_k} x_{ijk} = 1, \qquad \forall j \in \mathcal{I}_n, \; \forall k \in \mathcal{I}_p. \tag{2}$$

This is different from the assumption made by [4] and [17] where customer demand may be satisfied from a warehouse outside the time window. Since the demand of customer j for part k is a random variable with Poisson distribution, it follows that demand for part k experienced at warehouse i during lead time is a Poisson random variable $\widetilde{\lambda}_{ik}$ with mean

$$\lambda_{ik} = t_{ik} \sum_{j \in \mathcal{I}_n :\, \tau_{ijk} \le w_k} d_{jk} x_{ijk}. \tag{3}$$

Given the mean demand λ_{ik} and the stock level s_{ik} for part k at warehouse i, the achieved service level for part k at each warehouse is defined as $\beta(\lambda_{ik}, S_{ik}) = \Pr(\widetilde{\lambda}_{ik} \le S_{ik} - 1)$. This is known as the Type II service level or fill rate in the case of backorders [20]. Finally the target service levels are denoted by α_{ik} and α_k for part-warehouse specific, and part specific service levels, respectively.

2.1 Part-warehouse specific service levels

In the part-warehouse case, warehouse i is required to deliver the assigned demand of part k $100\alpha_{ik}\%$ of time within the time window w_k. This is achieved when "the probability that stock availability during lead time is strictly larger than demand" is at least α_{ik}. In other words, the stock level S_{ik} of part k at an open warehouse i satisfies

$$S_{ik} \leq S_{\max} y_i, \qquad \forall i \in \mathcal{I}_m, \quad \forall k \in \mathcal{I}_p. \tag{4}$$

$$\Pr(\widetilde{\lambda}_{ik} \leq S_{ik} - y_i) \geq \alpha_{ik}, \quad \forall i \in \mathcal{I}_m, \quad \forall k \in \mathcal{I}_p. \tag{5}$$

We now present a nonlinear model for the location-inventory problem with part-warehouse specific service levels, denoted by **SM**.

$$\mathbf{SM}: \quad \min \quad \sum_{i \in \mathcal{I}_m} f_i y_i + \sum_{(i,j,k) \in \mathcal{I}_m \times \mathcal{I}_n \times \mathcal{I}_p} c_{ijk} d_{jk} x_{ijk} + \sum_{(i,k) \in \mathcal{I}_m \times \mathcal{I}_p} h_{ik} S_{ik} \tag{6}$$

$$\text{s.t.} \quad \sum_{i \in \mathcal{I}_m :\, \tau_{ijk} \leq w_k} x_{ijk} = 1, \qquad \forall j \in \mathcal{I}_n, \ \forall k \in \mathcal{I}_p \tag{7}$$

$$0 \leq x_{ijk} \leq y_i, \qquad \forall i \in \mathcal{I}_m, \ \forall j \in \mathcal{I}_n, \ \forall k \in \mathcal{I}_p \tag{8}$$

$$S_{ik} \leq S_{\max} y_i, \qquad \forall i \in \mathcal{I}_m, \ \forall k \in \mathcal{I}_p \tag{9}$$

$$\beta(\lambda_{ik}, S_{ik}) = \Pr(\widetilde{\lambda}_{ik} \leq S_{ik} - y_i) \geq \alpha_{ik}, \qquad \forall i \in \mathcal{I}_m, \ \forall k \in \mathcal{I}_p \tag{10}$$

$$y_i \in \{0, 1\}, \qquad \forall i \in \mathcal{I}_m \tag{11}$$

$$S_{ik} \in \{0, 1, 2, \ldots, S_{\max}\}, \qquad \forall i \in \mathcal{I}_m \ \forall k \in \mathcal{I}_p. \tag{12}$$

The objective function (6) minimizes the total cost of opening warehouses, transportation, and inventory holding. Constraints (7) ensure that customer demand for part k is assigned to warehouses that are within the time window. Constraints (8) restrict customer allocation to open warehouses. Constraints (9) allow positive stock levels for open warehouses only. Constraints (10) are service level constraints and state that the stock level should be high enough so that the probability of the part available is larger than the target service level α_{ik}.

Constraints (10) are highly nonlinear and present the main source of complication in **SM**. By exploiting properties of the fill-rate function $\beta(\lambda_{ik}, S_{ik})$, we are able to reformulate **SM** as a linear mixed integer program. For ease of presentation, we drop the indices i and k in the rest of this section. Under Poisson demand, the function $\beta(\lambda, S)$ in (10) is given by

$$\beta(\lambda, S) = e^{-\lambda} \sum_{r=0}^{S-1} \frac{\lambda^r}{r!}, \qquad \lambda \in [0, \infty), S \in \{1, 2, \ldots\}\,. \tag{13}$$

and has the following properties. First, for all positive integer S the fill-rate equals 1 when the demand rate is 0, $\beta(0, S) = 1$. Second, when the demand rate tends to ∞, the fill-rate tends to 0, $\lim_{\lambda \to \infty} \beta(\lambda, S) = 0$. Finally, function β is strictly monotonically decreasing with respect to $\lambda \in (0, \infty)$ since

$$\frac{d\beta(\lambda, S)}{d\lambda} = -e^{-\lambda} \frac{\lambda^{S-1}}{(S-1)!} < 0. \tag{14}$$

In other words for a given stock S, fill rate decreases when demand rate increases. It follows that for any $\alpha \in (0, 1)$, the equation $\beta(\lambda, S) = \alpha$ has a unique solution denoted by $\lambda(S, \alpha)$. That is to say for a given stock S the service level α is achieved exactly for only one demand rate $\lambda(S, \alpha)$. Therefore, $\beta(\lambda, S)$ meets or exceeds the service level α for demand rate $\lambda(S, \alpha)$ or lower, i.e., $\beta(\lambda, S) \geq \alpha$ if and only if $\lambda \leq \lambda_S(\alpha)$. These results are illustrated in Figure 2 for $\alpha = 0.5, 0.9$ and $S = 1, 2, \ldots, 6$.

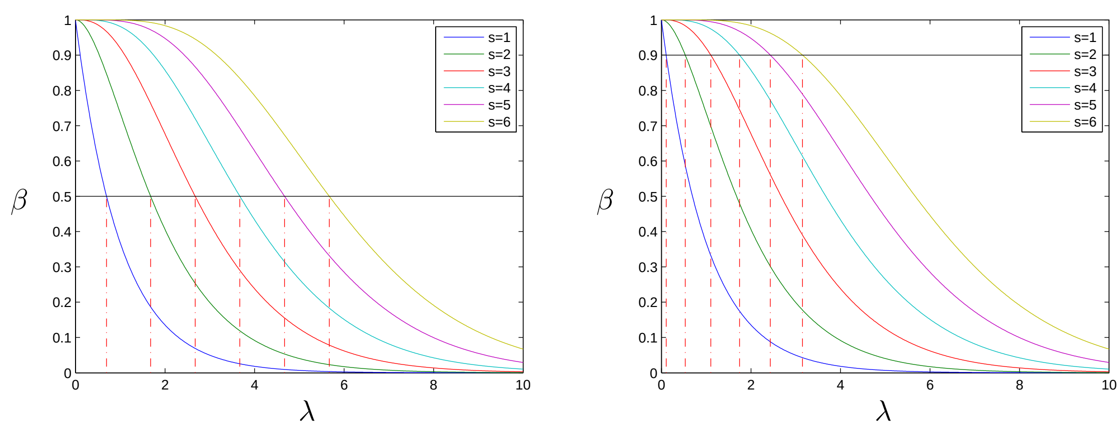

Figure 2: Graphs of $\beta(\lambda, S)$ with respect to λ for $S = 1, 2, \ldots, 6$. The λ values corresponding to dashed-dotted vertical lines are $\lambda_S(\alpha)$, the unique solutions of $\beta(\lambda, S) = \alpha$ for $\alpha = 0.5, 0.9$.

Based on the properties of the function β, we replace nonlinear constraints (9) and (10) by the following linear equivalents

$$\sum_{s\in\mathcal{I}_{S_{\max}}} V_{iks} \leq y_i, \qquad \forall i \in \mathcal{I}_m,\ \forall k \in \mathcal{I}_p, \tag{15}$$

$$\lambda_{ik} \leq \sum_{s\in\mathcal{I}_{S_{\max}}} \lambda_s(\alpha_{ik}) V_{iks}, \qquad \forall i \in \mathcal{I}_m\ \forall k \in \mathcal{I}_p. \tag{16}$$

where $\mathcal{I}_{S_{\max}} = \{1, 2, \ldots, S_{\max}\}$ and V_{iks} is a binary variable that takes value 1 if the base stock level for part k at warehouse i is $s \in \mathcal{I}_{S_{\max}}$, and 0 otherwise.

Replacing (9) and (10) by (15) and (16), we obtain an equivalent reformulation of the nonlinear mixed integer problem **SM** as a mixed integer linear program, denoted by **LMPW**.

$$\textbf{LMPW}: \quad \min \quad \sum_{i\in\mathcal{I}_m} f_i y_i + \sum_{(i,j,k)\in\mathcal{I}_m\times\mathcal{I}_n\times\mathcal{I}_p} c_{ijk} d_{jk} x_{ijk} + \sum_{(i,k,s)\in\mathcal{I}_m\times\mathcal{I}_p\times\mathcal{I}_{S_{\max}}} s h_{ik} V_{iks} \tag{17}$$

$$\text{s.t.} \quad \sum_{i\in\mathcal{I}_m:\ \tau_{ijk}\leq w_k} x_{ijk} = 1, \qquad \forall j \in \mathcal{I}_n,\ \forall k \in \mathcal{I}_p \tag{18}$$

$$0 \leq x_{ijk} \leq y_i, \qquad \forall i \in \mathcal{I}_m,\ \forall j \in \mathcal{I}_n,\ \forall k \in \mathcal{I}_p \tag{19}$$

$$\sum_{s\in\mathcal{I}_{S_{\max}}} V_{iks} \leq y_i, \qquad \forall i \in \mathcal{I}_m,\ \forall k \in \mathcal{I}_p \tag{20}$$

$$t_{ik} \sum_{j\in\mathcal{I}_n:\ \tau_{ijk}\leq w_k} d_{jk} x_{ijk} \leq \sum_{s\in\mathcal{I}_{S_{\max}}} \lambda_s(\alpha_{ik}) V_{iks}, \qquad \forall i \in \mathcal{I}_m, \forall k \in \mathcal{I}_p \tag{21}$$

$$y_i \in \{0,1\}, \qquad \forall i \in \mathcal{I}_m, \tag{22}$$

$$V_{iks} \in \{0,1\}, \qquad \forall i \in \mathcal{I}_m, \forall k \in \mathcal{I}_p, \forall s \in \mathcal{I}_{S_{\max}}. \tag{23}$$

In this model, the integer variables S_{ik} are removed and new binary variables V_{iks} are introduced. For any open warehouse i and any part k, there must exist only one s such that $V_{iks} = 1$, and for all closed warehouses we must have $V_{iks} = 0$; these are modeled by inequalities (20). The stock level of warehouse i for part k is given by $S_{ik} = \sum_{s\in\mathcal{I}_{S_{\max}}} sV_{iks}$. The values $\lambda_s(\alpha)$ are input data of **LMPW**, calculated as the unique solutions of $\beta(\lambda, s) = \alpha$ with respect to λ for given α and s using a Newton method.

2.2 Part specific service levels

In this scenario the manufacturer sets a system-wide target service level, which needs to be achieved by all warehouses collectively. A typical service requirement may read "demand for part k is satisfied in $100\alpha_k\%$ of the time within time window w_k." To model this requirement, we use a weighted average of service levels achieved at each warehouse to determine the overall service level. The service level constraints are expressed as:

$$\frac{\sum\limits_{(i,j)\in\mathcal{I}_m\times\mathcal{I}_n:\ \tau_{ijk}\le w_k} d_{jk}x_{ijk}\Pr(\widetilde{\lambda}_{ik}\le S_{ik}-1)}{\sum\limits_{j\in\mathcal{I}_n} d_{jk}} \ge \alpha_k, \qquad \forall k\in\mathcal{I}_p. \tag{24}$$

In the left hand side, the numerator is the sum of the service levels achieved at each warehouse $\Pr(\widetilde{\lambda}_{ik}\le S_{ik}-1)$ weighted by the demand assigned to the warehouse $d_{jk}x_{ijk}$. The denominator is the total demand rate for part k. Inequalities (24) allow individual part-warehouse service levels $\Pr(\widetilde{\lambda}_{ik}\le S_{ik}-1)$ to be higher or lower than the part target level α_k but guarantee that their weighted sum satisfies α_k. Therefore, they are relaxations of constraints (10) and provide more flexibility for the suppliers in fulfilling the service level requirements. A major disadvantage of these service level constraints is that they are highly nonlinear.

Using definition (3), we rewrite (24) as

$$\sum_{i\in\mathcal{I}_m}\frac{1}{t_{ik}}\lambda_{ik}\beta(\lambda_{ik},S_{ik}) \ge \alpha_k\sum_{j\in\mathcal{I}_n} d_{jk}, \qquad \forall k\in\mathcal{I}_p. \tag{25}$$

To linearize (25), we propose an approach that directly approximates the left-hand-side of the above constraints. This resolves all nonlinearities at once and leads to an effective reformulation as shown in the computational study. We define $\bar{z}_s = \max\limits_{\lambda}\lambda\beta(\lambda,s)$ for all $s\in\mathcal{I}_{S_{\max}}$. Let N be a positive integer number, which denotes the number of discretization points in the approximation. We define $z_{fs}=\frac{f}{N}\bar{z}_s$ for all $f\in\mathcal{I}_N=\{1,2,\ldots,N\}$ and denote the two roots of $\lambda\beta(\lambda,s)=z_{fs}$ by λ^1_{fs} and λ^2_{fs} where $\lambda^1_{fs}\le\lambda^2_{fs}$; see Figure 3 for an illustration.

Define variable Z_{ik} which equals z_{fs}, for some f and s; and binary variable V_{ikfs} which

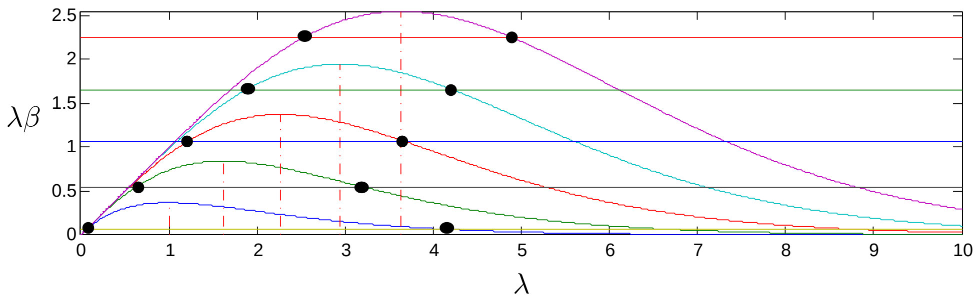

Figure 3: Graphs of $\lambda\beta(\lambda, s)$ with respect to λ for $s = 1, 2, \ldots, 5$. The function values corresponding to dashed-dotted vertical lines are $\bar{z}_s = \max_{\lambda} \lambda\beta(\lambda, s)$. The λ values corresponding to the dots are λ^1_{fs} and λ^2_{fs}, the two roots of $\lambda\beta(\lambda, s) = z_{fs}$ with $f = \bar{z}_s - 0.3$.

takes value 1 if $Z_{ik} = z_{fs}$. It follows that $Z_{ik} = \sum_{f,s} z_{fs} V_{ikfs}$. Using variables Z_{ik} to approximate $\lambda\beta(\lambda, s)$, the nonlinear constraints (25) are approximated by

$$\sum_{(i)\in(\mathcal{I}_m)} \frac{1}{t_{ik}} Z_{ik} \geq \alpha_k \sum_{j\in\mathcal{I}_n} d_{jk}, \qquad \forall k \in \mathcal{I}_p, \tag{26}$$

$$\lambda_{ik}\beta(\lambda_{ik}, S_{ik}) \geq Z_{ik}, \qquad \forall i \in \mathcal{I}_m, \forall k \in \mathcal{I}_p, \tag{27}$$

Now, replacing Z_{ik} by $\sum_{f,s} z_{fs} V_{ikfs}$ and using λ^1_{fs} and λ^2_{fs}, the two roots of $\lambda\beta(\lambda, s) = z_{fs}$, we rewrite (26) and (27) by the linear constraints

$$\sum_{(i,f,s)\in(\mathcal{I}_m\times\mathcal{I}_N\times\mathcal{I}_{S_{\max}})} \frac{1}{t_{ik}} z_{fs} V_{ikfs} \geq \alpha_k \sum_{j\in\mathcal{I}_n} d_{jk}, \qquad \forall k \in \mathcal{I}_p. \tag{28}$$

$$\sum_{(f,s)\in(\mathcal{I}_N\times\mathcal{I}_{S_{\max}})} \lambda^1_{fs} V_{ikfs} \leq \lambda_{ik} \leq \sum_{(f,s)\in(\mathcal{I}_N\times\mathcal{I}_{S_{\max}})} \lambda^2_{fs} V_{ikfs}, \qquad \forall i \in \mathcal{I}_m, \forall k \in \mathcal{I}_p, \tag{29}$$

The approximate linear model with part-specific service level constraints is denoted by **LMP** and is presented as

$$\textbf{LMP}: \quad \min \quad \sum_{i\in\mathcal{I}_m} f_i y_i + \sum_{(i,j,k)\in\mathcal{I}_m\times\mathcal{I}_n\times\mathcal{I}_p} c_{ijk} d_{jk} x_{ijk} + \sum_{(i,k,f,s)\in\mathcal{I}_m\times\mathcal{I}_p\times\mathcal{I}_N\times\mathcal{I}_{S_{\max}}} sh_{ik} V_{ikfs} \tag{30}$$

$$\text{s. t.} \quad \sum_{i\in\mathcal{I}_m:\, \tau_{ijk}\le w_k} x_{ijk} = 1, \qquad \forall j\in\mathcal{I}_n,\ \forall k\in\mathcal{I}_p \tag{31}$$

$$0 \le x_{ijk} \le y_i, \qquad \forall i\in\mathcal{I}_m,\ \forall j\in\mathcal{I}_n,\ \forall k\in\mathcal{I}_p \tag{32}$$

$$\sum_{(f,s)\in\mathcal{I}_N\times\mathcal{I}_{S_{\max}}} V_{ikfs} \le y_i, \qquad \forall i\in\mathcal{I}_m,\ \forall k\in\mathcal{I}_p \tag{33}$$

$$\sum_{(i,f,s)\in(\mathcal{I}_m\times\mathcal{I}_N\times\mathcal{I}_{S_{\max}})} \frac{1}{t_{ik}} z_{fs} V_{ikfs} \ge \alpha_k \sum_{j\in\mathcal{I}_n} d_{jk}, \qquad \forall k\in\mathcal{I}_p \tag{34}$$

$$t_{ik} \sum_{j\in\mathcal{I}_n:\, \tau_{ijk}\le w_{ik}} d_{jk} x_{ijk} \ge \sum_{(f,s)\in(\mathcal{I}_N\times\mathcal{I}_{S_{\max}})} \lambda^1_{fs} V_{ikfs}, \qquad \forall i\in\mathcal{I}_m, \forall k\in\mathcal{I}_p \tag{35}$$

$$t_{ik} \sum_{j\in\mathcal{I}_n:\, \tau_{ijk}\le w_k} d_{jk} x_{ijk} \le \sum_{(f,s)\in(\mathcal{I}_N\times\mathcal{I}_{S_{\max}})} \lambda^2_{fs} V_{ikfs}, \qquad \forall i\in\mathcal{I}_m, \forall k\in\mathcal{I}_p \tag{36}$$

$$y_i \in \{0,1\}, \qquad \forall i\in\mathcal{I}_m \tag{37}$$

$$V_{ikfs} \in \{0,1\}, \qquad \forall i\in\mathcal{I}_m\ \forall k\in\mathcal{I}_p\ \forall f\in\mathcal{I}_N\ \forall s\in\mathcal{I}_{S_{\max}}. \tag{38}$$

Any optimal solution of **LMP** is a feasible solution of the nonlinear integrated model with part-specific service level constraints (25). Similarly, as $N\to\infty$, optimal solutions of **LMP** converge to an optimal solution of the nonlinear model.

3 Computational Study

In this section, we discuss implementation issues as well as numerical results obtained by solving instances of the location-inventory problem with service levels using our models and those in the literature. We generate random instances and solve them using model **LMPW** under part-warehouse service levels, and model **LMP** under part service levels. We also solve the same instances using the the approximate model described in [4] and denoted by **CK**, and using the decoupled approach denoted by **DeC**. The approximate model **CK** uses a discretization of the fill rate functions at the warehouses. The decoupled approach **DeC** first solves a logistics network design (LND) problem to determine open warehouses and customer allocations, then uses this information to solve a pure inventory stocking (IS) problem.

Second, we run tests to investigate the effect of problem parameters on the location and inventory decisions. Finally, we compare our part-specific model **LMP** with the models in [17] referred to as **JKP**. It is important to note that the comparison with [17] is to be taken with caution because the underlying assumptions are different. We treat the case of backorders as in [4], while [17] treats the case of lost sales. Because the fill rate functions under backorders and lost sales are different [20], the approximations developed in this paper do not apply to the models in [17] and vice-versa. The comparison of **LMP** and **JKP** is useful to compare SPL systems operating under backorders and lost sales. The models analyzed are summarized in Table 2.

Table 2: List of models compared.

LMPW	Exact linear reformulation under part-warehouse service levels with backorders.
LMP	Approximate linear reformulation under part service levels with backorders.
CK	Approximate model under backorders given in [4].
DeC	Solution of the location problem followed by the inventory stocking problem.
JKP	Approximate approach under lost sales given in [17].

We used MATLAB version 7 computing software package and CPLEX version 11 optimization software package in our computational experiments. These applications were run on a machine with a 64bit Windows XP operating system, a 3GHz quad core Intel processor, and 8GBs of RAM. We devise a random generation procedure to generate instances to compare **LMPW**, **LMP**, **CK**, and **DeC** and to investigate the effect of problem parameters. Then, we use the same random instances from [17] to compare **LMP** and **JKP**.

3.1 Generating random data

The random instances are constructed by generating uniformly distributed random points for customer and warehouse locations in the unit square. The distance between a warehouse and a customer dist_{ij} is calculated as the Euclidean distance. The parameters are generated to reflect the characteristics of service parts, i.e., expensive parts and low demand. Inventory holding cost per unit ranges between 12.5% and 37.5% of part value v_k which is in turn uniformly distributed in $[2000, 3000]$. On the other hand the fixed cost is uniformly distributed in $[500, 1500]$. Holding cost is then relatively high compared to fixed costs. The time window w_k is 4 hours and the transportation time τ_{ijk} is 5 times the distance dist_{ij}. Since the distances are in the unit square, w_k and τ_{ijk} are within range and customers are not necessarily reachable from all locations. To ensure the instances are feasible, we check that every customer has at least one warehouse within the time window. Transportation cost c_{ijk} is generated in function of transportation time τ_{ijk}. Mean demand for part k γ_k for all customers is uniform in $[0, 1]$ and mean customer demand d_{jk} is Poisson distributed with mean γ_k. This leads to low demand levels. Finally, lead time t_{ik} is set to one week which is significantly higher than the time window w_k. Consequently, we expect to hold stock to satisfy demand during lead time and to achieve the target service levels. Parameter generation is detailed in the following list.

f_i	Uniform$[500, 1500]$	annual fixed cost at warehouse i,
v_k	Uniform$[2000, 3000]$	value of part k,
h_{ik}	Uniform$[0.125v_k, 0.375v_k]$	holding cost of part k at warehouse i,
τ_{ijk}	$5\,\text{dist}_{ij}$	transportation time between warehouse i and customer j,
c_{ijk}	$(a + b\tau_{ijk})$	transportation cost between warehouse i and customer j,
		where $a \sim$ Uniform$[100, 300]$ and $b \sim$ Uniform$[10, 20]$
γ_k	Uniform$[0, 1]$	mean customer demand for part k,
d_{jk}	Poisson$[\gamma_k]$	customer j's demand for part k,
w_k	4 hours	time window,
t_{ik}	1 week	lead time.

Table 3: Sizes of the solved instances.

Description	(m, n, p)	$S_{\max}$	N{**CK**,**LMP**}
Small-scale	$(5, 50, 3)$	5	{10,100}
Medium-scale	$(10, 100, 3)$	5	{10,100}
Large-scale	$(20, 300, 5)$	10	{10,10}
Single-part	$(50, 1000, 1)$	10	{10,10}

We generate four categories based on size with 10 instances in each category. The small-scale instances have 5 warehouses, 50 customers, and 3 parts. The medium-scale instances have 10 warehouses, 100 customers, and 3 parts. The large-scale instances have 20 warehouses, 300 customers, and 5 parts. The single-part instances have 50 warehouses and 1000 customers. As suggested in [4], we set $N = 10$ for the **CK** model. Although increasing N results in more accurate solutions, it takes longer time or fails to find a solution due to memory limitations. On the other hand, setting $N = 100$ was reasonable for **LMP** in small and medium instances to obtain more accurate solutions in a relatively short time. The maximum stock is set to 5 for the small and medium scale instances and to 10 for large-scale and single-part instances. The sizes of solved instances are given in Table 3 where $m, n, p, S_{\max}$, and N are the number of candidate warehouses, number of customers, number of parts, maximum number of units in stock at each warehouse, and number of discretization points used in the approximations, respectively. To make all the models comparable, we set $\alpha_{ik} = \alpha_k$, and choose α_k randomly from the set $\{0.5, 0.6, 0.7, 0.8, 0.9\}$.

While problem size in these types of formulations is not necessarily suggestive of the difficulty, we summarize the number of variables and constraints for each model in Table 4. It is observed that **LMP** has substantially more variables than **LMPW** and **CK**.

Table 4: Number of continuous and binary variables, and number of constraints of all models.

		Number of variables		Number of constraints
		continuous	binary	
DeC	LND	mnp	m	$p(1+n+mn)$
	IS	0	$mpS_{\max}$	$p(1+mn)$
LMPW		mnp	$m(1+pS_{\max})$	$p(2m+mn+n)$
LMP		mnp	$m(1+pNS_{\max})$	$p(1+3m+mn+n)$
CK		$mnp+mpNS_{\max}$	$m(1+2pN+pS_{\max})$	$p(1+n+m(1+n+3N+4NS_{\max}))$

3.2 Comparison of models under backorders

Tables 5, 6, and 7 report average statistics on the location and inventory decisions and CPU time. The tables display the average over 10 instances of the number of open warehouses (whouse), number of units in stock (stock), cost of location (loc), transportation (transp), and inventory holding (inv), and total cost (total). Performance measures in terms of solution quality (%diff) and CPU time (CPU) are also reported. The percentage difference in total cost between **LMP** and the other models %diff is computed as $100 * (\mathbf{X} - \mathbf{LMP})/\mathbf{LMP}$, where **X**=**DeC**, **LMPW**, and **CK**.

Table 5: Average number of open warehouses, total stock, and total cost of 10 small-scale problems.

	whouse	stock	loc	transp	inv	total	%diff	CPU	
DeC	1.5	6.2	1402.80	4519.75	4009.55	9932.11	2.22	0.09	sec
LMPW	1.6	6.3	1530.34	4575.20	3798.61	9904.15	1.15	0.27	sec
LMP	1.6	6.4	1530.34	4475.27	3835.55	9841.17	0	13.51	sec
CK	2.0	6.3	1887.02	4436.60	3777.89	10101.51	6.04	18.20	sec

Referring to Tables 5, 6, and 7, both **LMPW** and **LMP** were solved to optimality for all instances while **CK** was out of memory in 10 large-scale problems. **LMPW** and **LMP** always find feasible solutions with respect to service constraints, but **CK** found solutions with slight violations of service level constraints in 5 out of 10 medium-scale problems. As suggested in [4], pure inventory stocking problems must be solved to improve the actual service levels, which in turn increases total stock. In small-scale instances, the total cost obtained by **CK** is worse than that of **DeC**. Increasing the number of discretization points N in **CK** from 10 to 30 did not improve the total cost in these cases. **LMP** finds feasible solutions that improve over **DeC** by 2.22%, which suggests that **CK** found feasible solutions that are far from optimal (6.04% higher than **LMP**). In terms of solution time, **LMP** is substantially more difficult than **LMPW**, which in turn is more difficult than **DeC**, as expected. When it finds a solution, **CK** takes much longer time. This is due to the larger size and lower sparsity of the **CK** model compared to **LMP**, as shown in Table 8.

Table 6: Average number of open warehouses, total stock, and total cost of 10 medium-scale problems.

	whouse	stock	loc	transp	inv	total	%diff	CPU	
DeC	2.0	8.4	1495.69	6124.92	5416.66	13037.26	8.81	0.31	sec
LMPW	1.6	7.2	1268.08	6680.02	4337.42	12285.52	0.59	2.65	sec
LMP	1.8	7.6	1641.02	6555.58	4064.83	12261.43	0	2.65	min
CK	2.6	7.8	2030.88	6368.85	4419.14	12818.87	5.50	3.66	hr

Table 7: Average number of open warehouses, total stock, and total cost of 10 large-scale problems.

	whouse	stock	loc	transp	inv	total	%diff	CPU	
DeC	5.11	30.78	3726.44	20206.02	17866.50	41798.95	12.92	3	sec
LMPW	3.78	22.33	2988.23	24029.25	10565.83	37583.31	0.92	25	min
LMP	3.89	23.11	3221.39	23339.08	10656.13	37216.59	0	5	hr
CK	failed for all 10 large-scale problems with out-of-memory status.								

LMP provided on average 6.04% and 5.50% improvement over **CK** in small and medium cases, respectively. The savings over a decomposed approach are significant: 2.22, 8.81, and 12.92% for small, medium and large-scale instances, respectively. These numbers suggest that savings increase with size.

Theoretically, the optimal solution to the nonlinear model of the part-warehouse specific case **SM** is a feasible solution to the nonlinear model of the part specific case with constraints (25) when the part service level is enforced at every warehouse. Consequently, we expect **LMP** to give lower total cost than **LMPW** when both are solved to optimality. However, it is possible for the opposite to occur since **LMP** is an approximation of the part-specific case with constraints (25) while **LMPW** is equivalent to **SM**. We observed such behavior in three instances, which are excluded from the calculation of %Diff.

Table 8: Average size, sparsity, and number of binary variables of 10 large-scale problems.

		row	column	nonzero elements	% sparsity	binary
DeC	LND	26707.5	25227.5	75622.5	99.9888	20
	IS	30	194.4	388.8	93.2190	194.4
LMPW		26798.5	25444.2	101145.4	99.9852	345.7
LMP		27012.5	30746.1	148211.9	99.9821	5538.6
CK		60069.6	35470.3	4703068.7	99.7815	2440.8

Finally, we solved 10 large-scale single-part problems for benchmark purposes. The results are summarized in Table 9. **CK** stopped with out-of-memory status, while **LMPW** and **LMP** were solved to optimality. The total cost obtained by **LMP** is the lowest among all models. This comparison once more confirms that the number of open warehouses, total stock, and total cost on average are lower in the integrated models than in the decoupled one.

Table 9: Average number of open warehouses, total stock, and total cost of 10 large-scale single-part problems.

	whouse	stock	loc	transp	inv	total	%diff	CPU
DeC	2.9	6.6	1716.24	7178.23	5138.02	14032.50	19.64	5.63 sec
LMPW	1.9	5.4	1370.26	8167.66	2239.89	11777.81	1.24	4.32 min
LMP	1.9	5.4	1355.89	8083.71	2274.00	11713.61	0	7.12 min
CK	failed for all 10 large-scale single-part problems with out-of-memory status.							

3.3 Impact of service levels

In this part of testing, we use medium-scale instances to analyze the impact of service levels on the location and inventory decisions. The service levels are set to $\{0.5, 0.6, 0.7, 0.8, 0.9\}$ for **LMPW**, and are set to $\{0.7, 0.8\}$ for **LMP**. The results are summarized in Table 10.

Table 10: The impact of service levels on the average number of open warehouses, total stocks, and total costs of 10 medium-scale problems.

	α	whouse	stock	loc	transp	inv	total	CPU	
LMPW	0.5	1.8	5.7	1476.46	5611.74	3150.19	10238.40	2.70	sec
	0.6	1.8	6.2	1532.99	5602.77	3351.13	10486.89	2.21	sec
	0.7	1.6	7.3	1316.28	5827.82	4061.44	11205.54	3.12	sec
	0.8	1.7	8.8	1395.45	5882.24	4572.09	11849.79	4.08	sec
	0.9	1.8	10.2	1512.95	6135.91	5235.34	12884.19	5.23	sec
LMP	0.7	1.7	7.4	1379.84	5762.54	4053.05	11195.43	3.68	min
	0.8	1.6	8.4	1389.15	5626.56	4468.23	11483.94	3.81	min

The general trend in both models is that total cost increases as service level increases. This is largely due to increased stocks and inventory holding costs. While these results are expected, there is no such trend in location and transportation costs. As service level increases, the average number of open warehouses and fixed cost may increase or decrease, showing no particular trend. In fact, the prescribed location/allocation decisions could actually be quite different for different service levels, which suggests that inventory control structure can profoundly impact the network design decisions. The models seem to prescribe relocation of warehouses in addition to stock increases to respond to service level increases.

Generally, there is an increase in transportation costs when service levels increase, but it is not consistent. It is somewhat hard to predict how service levels impact the transportation cost because warehouse locations and service levels have complex interactions. One possible explanation is that increased service levels puts pressure to carry more inventory; perhaps to mitigate its effect, the optimal solution shifts to another less expensive location/allocation solution. Or, after a threshold a new warehouse is opened to mitigate the countering effects, etc. Consequently, making LND decisions and inventory control decisions separately is likely to lead to suboptimal solutions.

3.4 Comparison of models with backorders and lost sales

We carry out tests to investigate the impact of operating under backorders versus lost sales. We solve the single-part instances from [17] using model **LMP**, and compare in terms of system measures like total costs, stock levels, and network structure; as well as in terms of solution quality and computational time requirements. The single-part instances are built on a network with 15 warehouse locations, and 50 customers generated randomly on a grid of 150×150. The generation procedure is detailed in [17]. The inventory cost h_i is the same for all warehouses in each instance and is set to $\{1, 10, 20, 50, 100\}$. Transportation cost from warehouse i to customer j, c_{ij}, is found by rounding one-tenth of the Euclidean distance to a positive integer. The annual mean demand for each customer is generated uniformly within $(1, 3)$. Replenishment lead time is set to 1 week. The maximum stock level $S_{\max}$ is set to 5. The target service level α is set to one of $\{0.4, 0.6, 0.8\}$. The time window is set to 40.Fifteen instances are generated by varying the inventory cost h_i and the target service level α. Then, the 15 instances are replicated twice with the same parameter settings but with different random seeds to obtain instances $1-15$, $16-30$, and $31-45$. The results on the 45 instances are summarized in two tables where the fixed cost f_i is set to 0 in Table 11 and to 1000 in Table 12. In the tables, α^A stands for the actual achieved service level determined by the total number of open warehouses (whouse) and units in stock (stock). While the solution of **LMP** provides one objective function value which is an upper bound on the optimal objective value of the nonlinear model with constraints (25), **JKP** provides a heuristic solution with a lower bound (LB) and an upper bound (UB) obtained from the outer approximation approach developed in the paper. The lower bound is found by ignoring the demand of customers outside the time window. The upper bound is found by solving a series of models starting with a higher service level than the required α and decreasing it iteratively until the solution of the relaxed model satisfies α. The gap (gap) between the lower and upper bounds shows the quality of the heuristic solution and is calculated as $100\frac{UB-LB}{LB}$. The last column gives the percentage difference in cost denoted by (diff) and calculated as $100\frac{total-UB}{total}$.

Unlike [17] we require that every customer has at least one warehouse within the time window. Together with the fact that there are no lost sales, some of the instances turn out to be infeasible, i.e., some customers are not within the time window of any warehouse. To make these instances feasible, we set the time window so that the condition is satisfied. Another major difference is that we assume backorders while [17] assumes lost sales. These assumptions lead to different fill rate functions [20], and should be kept in mind when interpreting the results in Tables 11 and 12.

When comparing the number of open warehouses, the number of units in stock and the total cost in both tables, we note that the models under backorders and lost sales assumptions seem to lead to comparable decisions in most instances. Under zero fixed cost $f_i = 0$, **LMP** finds the same number of open warehouses and number of units in stock as **JKP** in about 69% of the instances, and slightly lower numbers in 29% of the instances. Under nonzero fixed cost $f_i = 1000$, **LMP** finds fewer units in stock in 87% of the instances. The total cost of **LMP** solutions are slightly cheaper that those of **JKP** with 0.35% and 1.03% under zero and nonzero fixed costs, respectively. Keeping in mind that both models do not guarantee optimality, the differences in cost and network structure are not significant. Intstead, the analysis implies that operating under backorders and lost sales lead to comparable systems. This is explained by the fact that under both assumptions, most of the demand is satisfied within the window as enforced by the high target service levels.

Computational times in Tables 11 and 12 indicate that instances become more difficult as α and h increase for both **LMP** and **JKP**. When $f_i = 0$, **LMP** and **JKP** use comparable time except for 3 instances where **LMP** uses significantly more time. On the other hand, for $f_i = 1000$, **LMP** uses consistently less time. Comparing Tables 11 and 12, **JKP** takes much longer time when $f_i = 1000$ than when $f_i = 0$. This suggests that instances with nonzero fixed cost become more difficult for **JKP**. This is not the case for **LMP** where all instances take similar or less time to solve when $f_i = 1000$ than when $f_i = 0$.

4 Conclusions

We presented two integrated inventory location models for a service parts logistics network design problem. To the best of our knowledge, this paper presents the most comprehensive treatment of the problem with backorder assumption to date. The models arise from two classes of service levels and their corresponding inventory control structures. Each of the cases, i.e., part-warehouse and part specific, relates to some important type of inventory control structure

Table 11: Results of (15×50) problem instances with zero fixed cost ($f_i = 0$).

Instance			**LMP**					**JKP**							
	h	α	*total*	α^A	*whouse*	*stock*	*time*	*LB*	*UB*	gap	α^A	*whouse*	*stock*	*time*	*diff*
1	1	0.4	232	0.806	11	11	1	232	232	0	0.806	12	12	0	0
2	1	0.6	232	0.803	12	12	0	232	232	0	0.806	12	12	0	0
3	1	0.8	232	0.813	12	12	1	232	232	0	0.806	12	12	0	0
4	10	0.4	314	0.734	7	7	0	314	314	0	0.738	7	7	0	0
5	10	0.6	314	0.733	7	7	1	314	314	0	0.735	7	7	0	0
6	10	0.8	327	0.812	8	9	2	321	321	0	0.802	9	9	12	1.83
7	20	0.4	368	0.588	4	4	1	369	369	0	0.654	5	5	0	-0.27
8	20	0.6	369	0.661	5	5	1	369	369	0	0.653	5	5	0	0
9	20	0.8	414	0.808	7	8	1	410	411	0.23	0.801	9	9	26	0.72
10	50	0.4	478	0.446	3	3	1	491	491	0	0.584	4	4	0	-2.72
11	50	0.6	519	0.662	5	5	4	491	519	5.32	0.654	5	5	18	0
12	50	0.8	654	0.808	7	8	2	650	652	0.35	0.805	7	8	82	0.31
13	100	0.4	628	0.446	3	3	1	659	665	0.8	0.414	3	3	140	-5.89
14	100	0.6	769	0.661	5	5	19	691	769	10.09	0.653	5	5	58	0
15	100	0.8	1054	0.808	7	8	13	1050	1052	0.22	0.803	7	8	127	0.19
16	1	0.4	279	0.775	15	15	1	279	279	0	0.778	15	15	0	0
17	1	0.6	279	0.78	15	15	0	279	279	0	0.776	15	15	0	0
18	1	0.8	281	0.83	15	17	6	280	280	0	0.802	15	16	0	0.36
19	10	0.4	363	0.714	8	8	1	368	368	0	0.737	9	9	0	-1.38
20	10	0.6	363	0.718	8	8	1	368	368	0	0.737	9	9	0	-1.38
21	10	0.8	391	0.808	9	11	4	385	388	0.81	0.806	9	11	15	0.77
22	20	0.4	438	0.59	6	6	1	443	443	0	0.651	7	7	0	-1.14
23	20	0.6	439	0.657	7	7	1	443	443	0	0.666	7	7	0	-0.91
24	20	0.8	501	0.809	9	11	46	489	490	0.21	0.8	8	10	30	2.20
25	50	0.4	564	0.451	4	4	5	583	583	0	0.43	4	4	0	-3.37
26	50	0.6	620	0.607	6	6	4	626	629	0.43	0.602	6	6	45	-1.45
27	50	0.8	819	0.807	7	10	1653	789	790	0.13	0.8	8	10	61	3.54
28	100	0.4	764	0.45	4	4	85	740	783	5.52	0.43	4	4	47	-2.49
29	100	0.6	920	0.605	6	6	3	926	929	0.29	0.6	6	6	139	-0.98
30	100	0.8	1319	0.807	7	10	109	1289	1290	0.08	0.8	8	10	122	2.20
31	1	0.4	290	0.684	11	11	1	290	290	0	0.679	11	11	0	0
32	1	0.6	290	0.683	11	11	0	290	290	0	0.683	11	11	0	0
33	1	0.8	299	0.81	11	20	6	296	298	0.67	0.802	11	19	35	0.33
34	10	0.4	366	0.617	6	6	1	366	366	0	0.605	6	6	0	0
35	10	0.6	366	0.614	6	6	1	366	366	0	0.605	6	6	0	0
36	10	0.8	441	0.805	7	14	8	427	438	2.52	0.801	7	14	159	0.68
37	20	0.4	414	0.486	4	4	1	416	416	0	0.541	5	5	0	-0.48
38	20	0.6	426	0.617	6	6	1	426	426	0	0.605	6	6	1	0
39	20	0.8	581	0.805	7	14	105	556	579	3.88	0.804	7	14	224	0.34
40	50	0.4	512	0.426	3	3	1	524	533	1.62	0.405	3	3	1	-4.10
41	50	0.6	606	0.612	6	6	2	588	606	2.86	0.606	6	6	35	0
42	50	0.8	1001	0.805	7	14	1952	916	999	8.25	0.804	7	14	269	0.20
43	100	0.4	662	0.426	3	3	3	674	683	1.26	0.405	3	3	9	-3.17
44	100	0.6	906	0.618	6	6	5	838	906	7.43	0.606	6	6	78	0
45	100	0.8	1701	0.805	7	14	3714	1516	1699	10.74	0.803	7	14	363	0.12

Table 12: Results of (15×50) problem instances with nonzero fixed cost ($f_i = 1000$).

Instance			**LMP**					**JKP**							
	h	α	*total*	α^A	*whouse*	*stock*	*time*	*LB*	*UB*	gap	α^A	*whouse*	*stock*	*time*	*diff*
1	1	0.4	2443	0.434	2	4	1	2443	2444	0.04	0.405	2	5	35	-0.04
2	1	0.6	3365	0.616	3	6	1	3365	3366	0.03	0.601	3	7	59	-0.03
3	1	0.8	4299	0.822	4	8	1	4299	4300	0.02	0.809	4	9	18	-0.02
4	10	0.4	2479	0.434	2	4	1	2479	2489	0.4	0.404	2	5	56	-0.40
5	10	0.6	3419	0.616	3	6	1	3419	3435	0.45	0.608	3	7	85	-0.47
6	10	0.8	4371	0.822	4	8	2	4371	4381	0.23	0.803	4	9	39	-0.23
7	20	0.4	2519	0.434	2	4	0	2519	2539	0.79	0.405	2	5	70	-0.79
8	20	0.6	3479	0.616	3	6	1	3479	3505	0.73	0.606	3	7	67	-0.75
9	20	0.8	4451	0.822	4	8	1	4451	4471	0.45	0.803	4	9	54	-0.45
10	50	0.4	2633	0.402	2	3	1	2639	2689	1.89	0.405	2	5	70	-2.13
11	50	0.6	3659	0.616	3	6	2	3659	3715	1.51	0.606	3	7	55	-1.53
12	50	0.8	4691	0.822	4	8	1	4691	4741	1.07	0.809	4	9	47	-1.07
13	100	0.4	2783	0.402	2	3	0	2824	3024	7.08	0.417	2	5	86	-8.66
14	100	0.6	3959	0.616	3	6	1	3959	4065	2.66	0.606	3	7	161	-2.68
15	100	0.8	5091	0.822	4	8	1	5091	5191	1.96	0.809	4	9	65	-1.96
16	1	0.4	2534	0.438	2	4	1	2533	2535	0.08	0.409	2	5	36	-0.04
17	1	0.6	3448	0.612	3	8	1	3447	3449	0.06	0.605	3	9	10	-0.03
18	1	0.8	6343	0.807	6	17	12	6340	6342	0.03	0.805	6	16	210	0.02
19	10	0.4	2570	0.44	2	4	1	2560	2580	0.78	0.409	2	5	59	-0.39
20	10	0.6	3520	0.612	3	8	1	3510	3530	0.57	0.609	3	9	19	-0.28
21	10	0.8	6464	0.815	6	11	8	6454	6464	0.15	0.8	6	13	277	0
22	20	0.4	2610	0.44	2	4	1	2590	2630	1.54	0.409	2	5	64	-0.77
23	20	0.6	3600	0.611	3	8	1	3580	3620	1.12	0.605	3	9	28	-0.56
24	20	0.8	6574	0.809	6	11	4	6574	6594	0.3	0.803	6	12	806	-0.30
25	50	0.4	2730	0.438	2	4	1	2680	2780	3.73	0.415	2	5	80	-1.83
26	50	0.6	3840	0.612	3	8	1	3790	3890	2.64	0.609	3	9	57	-1.30
27	50	0.8	6904	0.815	6	11	7	6934	6954	0.28	0.8	6	12	422	-0.72
28	100	0.4	2930	0.442	2	4	3	2830	3030	7.07	0.403	2	5	54	-3.41
29	100	0.6	4240	0.612	3	8	1	4140	4340	4.83	0.609	3	9	74	-2.36
30	100	0.8	7454	0.814	6	11	4	7534	7554	0.26	0.803	6	12	428	-1.34
31	1	0.4	2460	0.427	2	4	1	2460	2462	0.08	0.414	2	6	12	-0.08
32	1	0.6	3381	0.607	3	7	1	3383	3385	0.06	0.601	3	11	8	-0.12
33	1	0.8	6330	0.804	6	20	5	6325	6326	0.02	0.8	6	16	40	0.06
34	10	0.4	2496	0.427	2	4	1	2496	2530	1.35	0.421	2	5	23	-1.36
35	10	0.6	3443	0.612	3	6	1	3443	3453	0.29	0.602	3	7	12	-0.29
36	10	0.8	6493	0.803	6	18	22	6460	6470	0.15	0.8	6	16	87	0.35
37	20	0.4	2536	0.427	2	4	1	2536	2580	1.73	0.421	2	5	23	-1.74
38	20	0.6	3503	0.612	3	6	1	3503	3523	0.57	0.607	3	7	11	-0.57
39	20	0.8	6666	0.802	6	17	47	6610	6630	0.3	0.8	6	16	126	0.54
40	50	0.4	2656	0.427	2	4	1	2656	2730	2.78	0.421	2	5	39	-2.79
41	50	0.6	3683	0.612	3	6	1	3683	3733	1.36	0.601	3	7	19	-1.36
42	50	0.8	7176	0.802	6	17	28	7060	7110	0.71	0.8	6	16	139	0.92
43	100	0.4	2856	0.427	2	4	1	2856	2980	4.33	0.421	2	5	40	-4.34
44	100	0.6	3983	0.612	3	6	1	3983	4083	2.51	0.604	3	7	30	-2.51
45	100	0.8	8026	0.802	6	17	50	7810	7910	1.29	0.801	6	16	211	1.45

that can widely be observed in practice. Regardless of the environment, our results provide a further justification for an integrated approach to network design, which can greatly improve the system cost over a two-step approach. Furthermore, the results suggest that our approach is more effective than those reported in the literature both in terms of solution quality and computational time.

Although our approach is particularly well-suited to the service parts environment, it could have wider applicability in other settings that involve difficult nonlinear constraints. In addition, our work can be extended in many ways. Some of the assumptions made may be relaxed, which would lead to more realistic models. For example, one may include multiple distribution centers with limited inventories, different service level measures, and part bundling or consolidation. Perhaps among the most important and difficult extensions is accurate modeling of environments where inventories are controlled centrally or collaboratively, which potentially involves rationing policies and emergency lateral shipments among warehouses. Finally, all approaches in the literature including this work use the ideas of outer or inner approximations and linearization of the nonlinear service constraints. An alternative promising approach is the use of decomposition methods like Lagrangian relaxation and Benders decomposition to decompose the models possibly into LND and inventory decisions.

References

[1] Cohen, M. A., Zheng, Y.-S., and Agrawal, V. (1997) Service parts logistics: a benchmark analysis. *IIE Transactions*, **29**, 627–639.

[2] Poole, K. (2003) Seizing the potential of the service supply chain. *Supply Chain Management Review*, **7**, 54–61.

[3] Kim, S. H., Cohen, M. A., and Netessine, S. (2009) Performance contracting in after-sales service supply chains. *Management Science*, **11**, 109–127.

[4] Candas, M. and Kutanoglu, E. (2007) Benefits of considering inventory in service parts logistics network design problems with time-based service constraints. *IIE Transactions*, **39**, 159–176.

[5] Baumol, W. J. and Wolfe, P. (1958) A warehouse-location problem. *Operations Research*, **6**, 252–263.

[6] Nozick, L. K. and Turnquist, M. A. (2001) Inventory, transportation, service quality and the location of distribution centers. *European Journal of Operational Research*, **129**, 362–371.

[7] Nozick, L. K. and Turnquist, M. A. (2001) A two-echelon inventory allocation and distribution center location analysis. *Transportation Research Part E*, **37**, 421–441.

[8] Barahona, F. and Jensen, D. (1998) Plant location with minimum inventory. *Mathematical Programming*, **83**, 101–111.

[9] Erlebacher, S. J. and Meller, R. D. (2000) The interaction of location and inventory in designing distribution systems. *IIE Transactions*, **32**, 155–166.

[10] Teo, C.-P., Ou, J., and Goh, M. (2001) Impact on inventory costs with consolidation of distribution centers. *IIE Transactions*, **33**, 99–110.

[11] Miranda, P. A. and Garrido, R. A. (2004) Incorporating inventory control decisions into a strategic distribution network design model with stochastic demand. *Transportation Research Part E: Logistics and Transportation Review*, **40**, 183–207.

[12] Miranda, P. A. and Garrido, R. A. (2006) A simultaneous inventory control and facility location model with stochastic capacity constraints. *Networks and Spatial Economics*, **6**, 39–53.

[13] Shen, Z.-J. M. (2000) *Efficient Algorithms for Various Supply Chain Problems*. Ph.D. Dissertation, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL 60208.

[14] Shen, Z.-J. M., Coullard, C., and Daskin, M. S. (2003) A joint location-inventory model. *Transportation Science*, **37**, 40–55.

[15] Shu, J., Teo, C.-P., and Shen, Z.-J. M. (2005) Stochastic transportation-inventory network design problem. *Operations Research*, **53**, 48–60.

[16] Shen, Z.-J. M. and Daskin, M. S. (2005) Trade-offs between customer service and cost in integrated supply chain design. *Manufacturing and Service Operations Management*, **7**, 188–207.

[17] Jeet, V., Kutanoglu, E., and Partani, A. (2009) Logistics network design with inventory stocking for low-demand parts: Modeling and optimization. *IIE Transactions*, **41**, 389–407.

[18] Sen, A., Bhatia, D., and Dogan, K. (2010) Applied materials uses operations research to design its service and parts network. *Interfaces*, **4**, 253–266.

[19] Mak, H.-Y. and Shen, Z.-J. M. (2009) A two-echelon inventory-location problem with service considerations. *Navel Research Logistics*, **56**, 601–609.

[20] Zipkin, P. (2000) Fundamentals of inventory management. *McGraw Hill-Irwin, Boston, MA*.